https://doi.org/10.20965/jrm.2021.p0537

Paper:

Investigation of Preliminary Motions from a Static State and Their Predictability

Chaoshun Xu*, **Masahiro Fujiwara****, **Yasutoshi Makino****,***, **and Hiroyuki Shinoda****

*Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
E-mail: xu@hapis.k.u-tokyo.ac.jp
**Graduate School of Frontier Sciences, The University of Tokyo
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan
***JST PRESTO
7 Gobancho, Chiyoda-ku, Tokyo 102-0076, Japan

[Received December 22, 2020; accepted March 30, 2021]

Humans observe the actions of others and predict their movements slightly ahead of time in everyday life. Many studies have been conducted to automate such a prediction ability computationally using neural networks; however, they implicitly assumed that preliminary motions occurred before significant movements. In this study, we quantitatively investigate when and how long a preliminary motion appears in motions from static states and what kinds of motion can be predicted in principle. We consider this knowledge fundamental for movement prediction in interaction techniques. We examined preliminary motions of basic movements such as kicking and jumping, and confirmed the presence of preliminary motions by using them as inputs to a neural network. As a result, although we did not find preliminary motion for a hand-moving task, a left-right jumping task had the most preliminary motion, up to 0.4 s before the main movement.

Keywords: motion capture, movement prediction, machine learning, robots in society

1. Introduction

Interactions between humans and machines are expected to increase in parallel with the accelerating development of robotics technology [1]. In order for robots to behave appropriately in human society, they must interact with humans [2, 3] and focus on observing safety protocols [4]. For this purpose, it is considered effective to observe and predict human movements. For example, long-term and macro-level prediction [5] of the behavior or flow of people within crowds can be applied to choose efficient movements for robots inside a group of people. On the other hand, short-term and personal-level prediction [6] of an individual's movements can be applied to tasks such as avoiding collisions in emergency situations. Rapid, subtle personal observations and predictions have also been used to guide the behavior of individuals in groups [7–9]. According to Ballerini et al. [8], individuals in a flock make behavioral decisions based on the movements of a few nearest neighbors. In addition, according to Lukeman et al. [9], the behavior of a herd can be represented by a rule that when a member's neighbor moves, that member also moves. Thus, mutual observation and prediction among individuals are thought to influence the behavior of groups. In this study, we aim to investigate the predictability of short-term and low-level (i.e., concerning a specific individual) predictions from the perspective of preliminary motion.

Many studies have been conducted to simulate human short-term movement prediction ability via calculations and computations [6, 10–16]. For example, Martinez et al. [6], Fragkiadaki et al. [10], and Chiu et al. [11] used regressive neural networks to predict future movements using the Human3.6M [12] dataset. Barsoum et al. [13] used generative adversarial networks to predict the future poses. Horiuchi et al. [16] used a forward-propagating neural network to predict movements for data on human jumping motions measured with Kinect. Previous studies have shown that movement prediction can be used in computationally augmented sport tasks, such as predicting the trajectory of a volleyball toss [17], the landing point of ping-pong ball serves [18], or future movements in martial arts [19].

Although many studies have been conducted on predicting human motions, few have quantitatively investigated when and how much of the preliminary motions had to be observed before a given motion can be predicted. It is expected that the time conditions for prediction will differ for movements involving shifting one's weight (e.g., walking and jumping) and those involving only hand movements (e.g., sign language). This is expected because all the muscles in the body must move in coordination to move a large mass in body movements involving weight shifting, and various preliminary movements are thus expected to occur. The same may occur to maintain balance in a situation where the center of gravity is not shifted, but the limbs are moved significantly. Clarifying the conditions for the occurrence of preliminary motions is expected to be important in applying prediction technology to the fields of human-robot and human-computer interaction.

We focus on motions performed from static states to explore the limits of automated prediction of human motion. This is because it remains difficult to distinguish in the case of successive motions between the end of one motion and the beginning of the next. These two motions appear simultaneously when the motion transitions from one state to the next. For example, when we consider a motion to take two steps forward and kick a ball, we cannot distinguish which part of the motion is the cessation of walking, and which part is a preliminary motion for kicking, at the moment of the transition between the two. In this study, we define a preliminary motion as motion preceding a target motion, which began from a static state, and analyzed these transitions. The purpose of this study is to identify limitations in the effectiveness of prediction methods for a single action and evaluate them in terms of prediction accuracy in NNs.

In this study, we first measured human motion beginning from a static state at 100 fps using a motion capture system (OptiTrack). For these motions, we looked at the movements of each body part and checked whether they moved prior to the start of the target movement. Depending on the type of movement, we identified several types of motion group, including those in which preliminary motions occurred and those in which they did not; among the groups in which preliminary motions occurred, we further distinguished whether the timing of their occurrence was relatively early or immediately prior to the primary motion. We applied these preliminary motions to a simple forward-propagating neural network to determine whether they could be used to predict subsequent motions.

As a result, we confirmed that there was no preliminary motion for hand waving movements, and that they were not predictable. In contrast, for left and right jumps, which were the focus of a previous study [16], preliminary motions occurred approximately 0.4 s prior to the main movement, and it was confirmed that the preliminary movement could be used to predict primary movements occurring in the near future. This result is consistent with that of the abovementioned previous study.

The contributions of this paper are as follows.

1) We identified a set of motions in which body parts began moving before the main motion, by observing body motion with respect to the starting point of the main motion.

2) We confirmed that for the group of motions with associated preliminary motions, we were able to use neural networks to predict future motions.

3) We demonstrate that we could not predict motions without such preliminary motions, showing the limitations of movement prediction based on preliminary motion.

This fundamental information about the presence or absence of preliminary motions can be used to predict instantaneous actions of people and determine the next actions of robots in environments where people and robots coexist. Prediction using high-speed vision has also been performed [20]; however, although high-speed machine vision can accurately predict information as long as relevant information is available, NN prediction approaches have the advantage of being able to predict next actions even in environments with latency. Therefore, such methods are desirable to compensate for communication delays in remote communication, such as in telexistence [21] or telepresence applications. Our results show the limits of the types of actions that can be predicted and are expected to be useful for the design of such systems.

2. Type of Motions Measured

2.1. Predictability of Body Movements

In terms of human body movements, some motions are more likely be associated with predictable preliminary motions, whereas others are less so or not at all. In general, the larger the movement, the greater the mass and velocity involved, and more force would be required, implying a probability of characteristic preparatory movements. According to Horiuchi et al. [16], movements of the body's limbs and extremities are more difficult to predict, whereas whole-body or trunk movements and shifts in center of mass or gravity are easier to predict. Examples of the former include movements of the arms and legs, and examples of the latter include movements such as jumping and walking.

If we can classify types of behavioral actions as predictable or unpredictable and quantitatively show how far ahead an automated system can predict, we can use this information to model group behavior among people who mutually predict each other. For example, walking actions require a shift in one's center of gravity, and walking is an easily predictable action, as shown in another study [22]. Therefore, it may become possible to simulate the behavior of people in crowds, such as in scramble intersections or other public spaces, including aspects of how well people can predict the behavior of others.

2.2. Classification of Body Movements

Movements of extremities are the only movements that can be performed by humans without relatively large center-of-gravity shifts. Accordingly, we considered three types of movements, including hand, foot, and neck movements, among which hand and foot movements were determined to be the most important. Specifically, we measured the following types of movement (**Fig. 1**).

- Raising the dominant arm and touching a virtual button at an angle of 45° to the left and right, approximately 30 cm in front and 30 cm to the left and right.

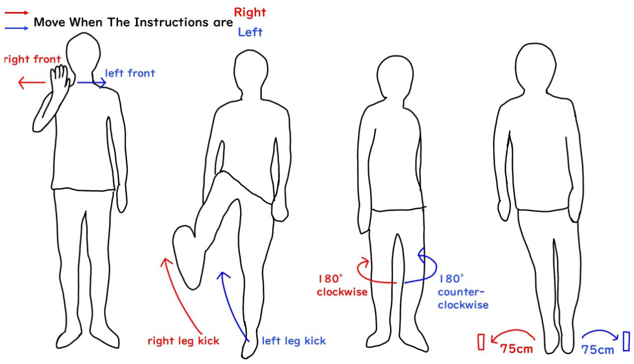

Fig. 1. Type of movements. From left, button touch, kick, rotation jump, and left-right jump.

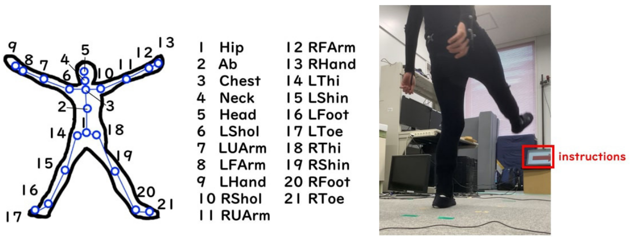

Fig. 2. The 21 body points measured with OptiTrack, and the actual motion capture scene.

- Kicking a right or left leg up approximately 60° with the leg extended.

In contrast, humans are capable of various full-body movements, such as jumping and walking. In this experiment, we decided to concentrate on jumping because we focused on movement from a stationary state. We also aimed to measure preliminary movements of rotation and translation preceding jumping motions. Specifically, we measured the following types of movement (**Fig. 1**).

- Jumping 180° rotating clockwise or counterclockwise while remaining in the same place.

- Jumping 75 cm to the left or right.

To summarize the above, we measured several types of motions such as arm movements, leg movements, and whole-body movements, which are relatively large motions in human motions, and are also thought to be important in human interaction. These motions were measured using an OptiTrack motion capture system and recorded as skeletal coordinates of 21 body points (**Fig. 2**).

Seven subjects (seven right-handed men in their 20s) participated in the motion capture procedures. For each movement, the participants were instructed to move as naturally and as quickly as possible. The subjects were instructed to move in the corresponding direction (left or right) shown on a monitor. For all subjects, 120 instances were recorded for each movement, collected at 100 fps. Instructions for the left and right directions were presented randomly. The initial position of the experiment and the target position for the jump were marked on the floor.

3. Experiment

3.1. Moment of the Preliminary Motions

When performing the intended main motions, some motions are performed unconsciously for reasons such as balance. Among these motions, we refer to those that began before the main motion as preliminary motions in the context of a static starting position. By observing the preliminary motion, the action of the main motion can be predicted before they begin. The goal of this study is to identify the timing of this preliminary motion and its location. Therefore, we first need to define the starting moment of the main action. Because the static state does not correspond to a preliminary motion, frames which may have a higher likelihood of being a preliminary motion can be identified by detecting a moment when each body part starts to move.

The starting point of the target motion should be the moment when we can determine the direction of movement of the subject is going to move by looking at the target body point. We set the starting point of the main motion as the point where the hands or feet moved 3 mm or more in the direction indicated for three consecutive frames in the button touch and kick movements. The threshold of 3 mm over 3 frames was determined by considering the noise motion during the static state (up to 0.5 mm/frame) and OptiTrack motion capture system accuracy (0.5 mm) to robustly detect the main movements. For the two jumping motions, we considered jumping to be "an action in which the feet leave the ground and the center of gravity moves upward from the initial state." Therefore, we considered that "the moment when the center of gravity begins to move upward" was an appropriate starting point to identify a jumping motion.

To determine the motion of all body parts starting to move from the static state, the mean and standard deviation of the velocity at static states were calculated for every body point as a noise component under static conditions. Then, a velocity threshold was empirically defined as the sum of three times the mean and standard deviation. The time when a velocity exceeded the threshold for three consecutive frames was set as the moment of initial movement of the various body parts.

3.2. Time and Duration of Preliminary Motions

Using the starting point of a target motion as a reference point, we varied the following two values as inputs to the NN for movement prediction.

1. Number of frames ahead used to predict (frames before movement, FBM).

2. Number of frames used as input (used frames, UF).

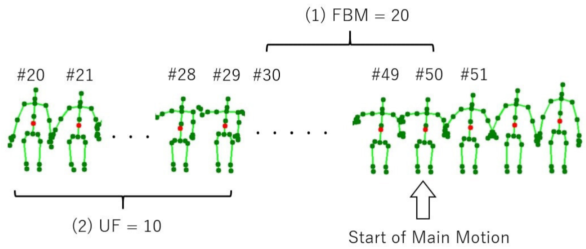

Fig. 3. Input of skeleton model. We use sequential frames of UF as an input.

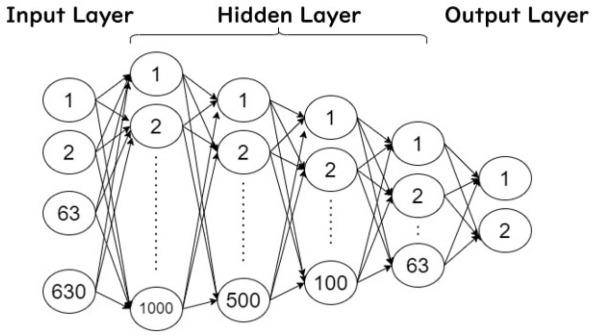

Fig. 4. The simple fully connected NN which Horiuchi et al. [16] used.

For example, if the target motion start point was at the 50th frame, and FBM was 20 and UF was 10, the data at frames 20–29 were used as input (**Fig. 3**).

For the NN, we used a fully connected network with four hidden layers, similar that of a previous study [16] as shown in **Fig. 4**. The input data were 3D coordinates of 21 joints measured using an OptiTrack motion capture system. The total number of units in the input layer was (21 joints) $\times$ (3 dimensions) $\times$ (# of UF). **Fig. 4** shows an example when the UF is 10. The middle layer was fixed at 1000, 500, 100, and 63 units, in that order, and the output layer had 2 (left or right). The activation function of each unit in the intermediate layer was a normalized linear function, and the output layer was an identity map. A softmax cross-entropy error function was used, as well as an Adam optimizer.

Four randomly chosen subjects out of seven were used as training data, two were used as validation data, and one as testing data. The training was conducted in mini-batches, with 100 minibatches and 1000 epochs.

4. Results

4.1. Moment of the Preliminary Motions

In **Figs. 5–8** below, the button touch movement is denoted by BT_Rhand, the kick movement is denoted by Kick_F, the rotation jump movement is denoted by Jump_Rot_180, and the left-right jump movement is denoted by JumpLR. The starting points of these target movements (mean $\pm$ standard deviation) and initial motion moments of the other body parts are summarized in **Table 1**. The column "Main motion" represents the time from the instruction to the starting point of the target motion. The column "Mean of motion start frame of other body parts" represents the average value of the difference between the starting point of the target motion and that of each body part. That is, when this value is negative, the other body parts tend to start moving prior to the main movement.

The average movement speed of the body parts in each frame based on the starting point is shown in **Figs. 5–8**. The body parts with larger movements are placed on the upper side, and those with smaller movements on the lower side. These figures were constructed only from right-side movement data, as the left-side movement data are symmetrical and largely equivalent.

These figures indicate that almost no preliminary motions occurred in the button touch movements, and after the initiation of arm motions, other body parts moved to maintain overall balance. The kick movement showed an average of 10 frames, and a maximum of 20–30 frames of preliminary motion. Before a kick, the hand on the opposite side of the moving foot started to move slightly earlier. The jumping movements showed an average of 20 frames and a maximum of 30–40 frames of preliminary motion before a jump began. In the rotation jump, a hand tends to move in the opposite direction as the jump just prior, while in the left-right jumps, a hand moved in the same direction as the jump prior to the action.

4.2. Time and Duration of Preliminary Motions

Having identified tendencies of possible preliminary motions, we determined the period of the input frames based on these frames. For the kicking and jumping motions, the number of used frames (UF) varied from 5 to 40 (0.05 to 0.40 s), and the frames before movement (FBM) varied from 0 to 40 frames before the start of the main motion. Because it was unlikely that a preliminary motion would be available for the button touch motion, the UF was 5 to 30 (0.05 to 0.30 s), and FBM was -20 to 10 (minus means the input is after the start of the main motion).

Using these data as input, the accuracy results of the left-right classification task for 120 test data (data of one subject not used for training) are summarized in **Tables 2–5**. We chose testing data from two different sets; one set used data from Subjects 1–4 as training data, from Subjects 5 and 6 as validation data, and that of Subject 7 as testing data, while the other set used data of Subjects 4–7 as training data, that of Subjects 2 and 3 as validation data, and that of Subject 1 as testing data. Because there was no significant difference in accuracy for UF values above 25, we omitted some of the calculations (shown in blank in the tables).

According to these tables, we may observe that for the button-touch movement, using a body movement 0.2 s "after" the motion started enabled prediction. There

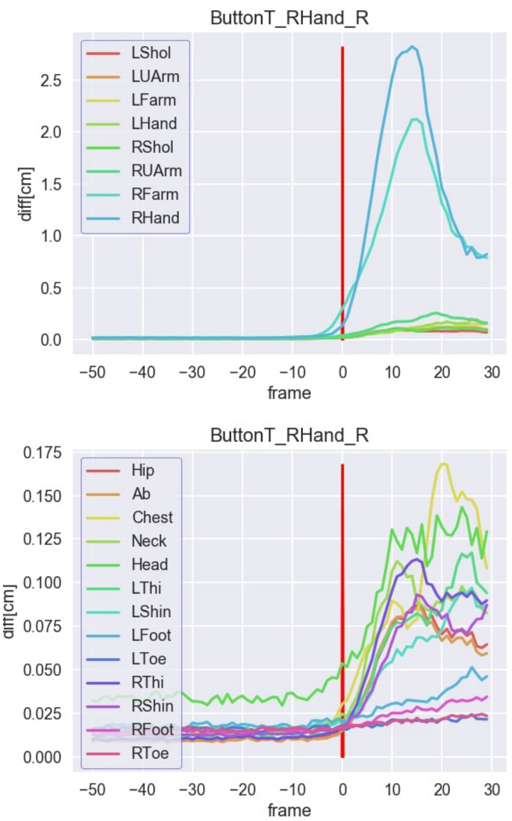

Fig. 5. Motion of body parts in the button touch movement.

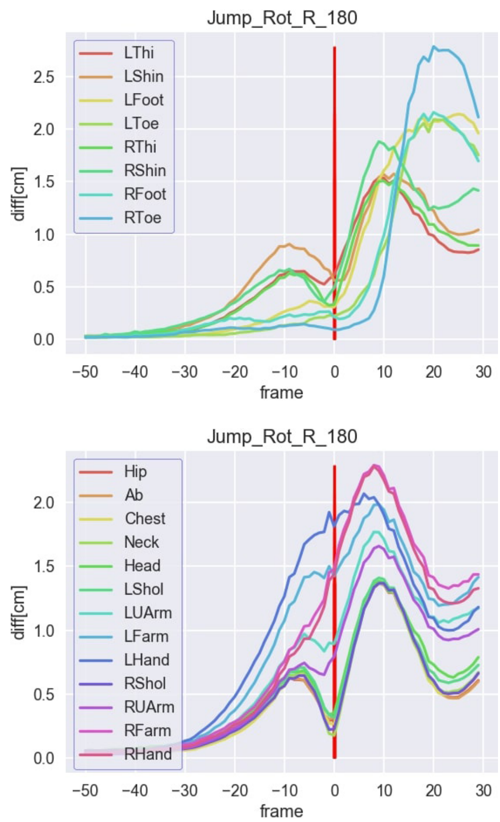

Fig. 7. Motion of body parts in the rotation jump movement.

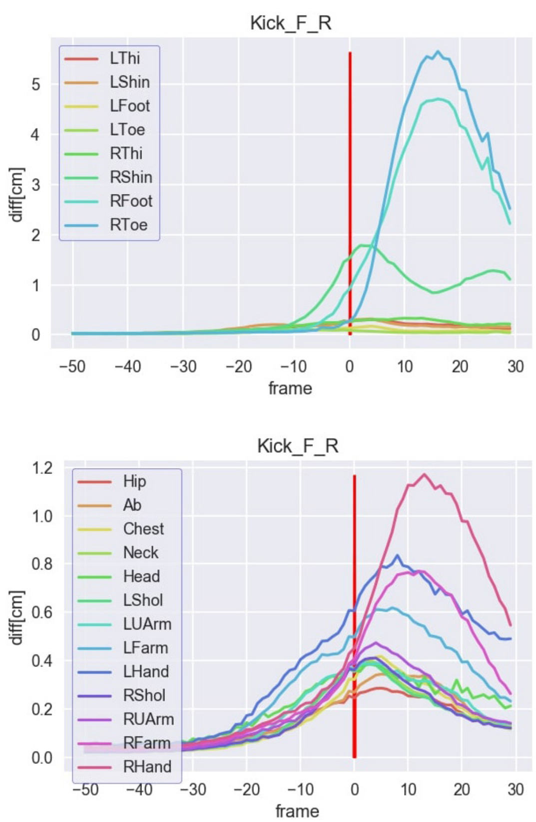

Fig. 6. Motion of body parts in the kick movement.

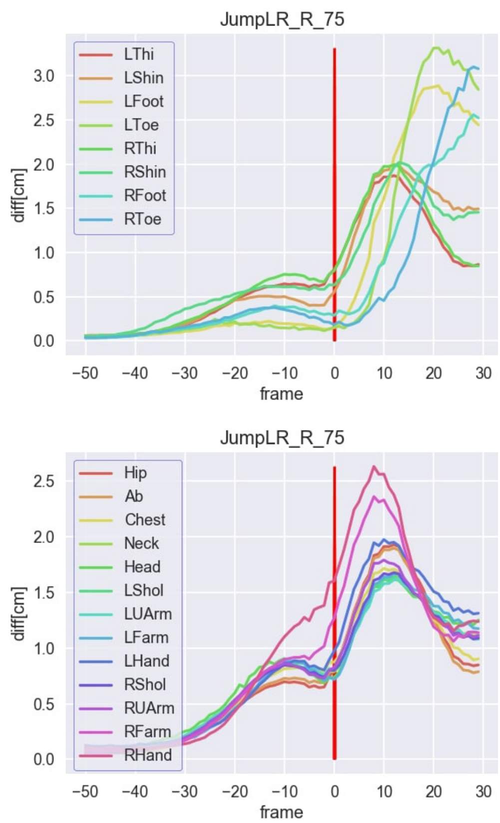

Fig. 8. Motion of body parts in the left-right jump movement.

Table 1. Start of the main motion and preliminary motion.

No.	Main motion [ms] (Mean SD), $n = 840$	Mean of motion start frame of other body parts [ms]
Button touch	379±65	+82
Kick	609±115	−109
Rotation jump	660±123	−200
Left-right jump	733±72	−221

Table 2. Classification accuracy on button touch movement.

(a) Subject 7 as testing data.

	FBM [100 fps]			
UF	−20	−10	0	10
5	1.00	0.83	0.47	0.48
10	1.00	0.88	0.47	0.52
15	1.00	0.76	0.47	0.48
20	1.00	0.63	0.47	0.49
25	1.00	0.78		
30	0.79	0.51		

(b) Subject 1 as testing data.

	FBM [100 fps]			
UF	−20	−10	0	10
5	0.98	0.82	0.48	0.49
10	0.97	0.83	0.47	0.51
15	0.95	0.73	0.52	0.50
20	0.95	0.67	0.52	0.54

Table 3. Classification accuracy on kick movement.

(a) Subject 7 as testing data.

	FBM [100 fps]				
UF	0	10	20	30	40
5	0.76	0.57	0.44	0.44	0.47
10	0.88	0.62	0.46	0.43	0.51
15	0.89	0.62	0.43	0.46	0.50
20	0.91	0.66	0.45	0.46	0.51
25	0.88	0.62			
30	0.84	0.61			
35	0.87	0.60			
40	0.85	0.58			

(b) Subject 1 as testing data.

	FBM [100 fps]				
UF	0	10	20	30	40
5	0.95	0.69	0.73	0.49	0.46
10	0.96	0.78	0.66	0.51	0.50
15	0.82	0.86	0.62	0.47	0.48
20	0.91	0.88	0.61	0.48	0.51

Table 4. Classification accuracy on rotation jump movement.

(a) Subject 7 as testing data.

	FBM [100 fps]				
UF	0	10	20	30	40
5	0.97	0.88	0.51	0.51	0.46
10	0.98	0.91	0.47	0.48	0.44
15	0.97	0.90	0.53	0.53	0.49
20	0.99	0.81	0.55	0.53	0.56
25	0.99	0.86	0.53		
30	0.99	0.91	0.46		
35	0.98	0.90	0.51		
40	0.99	0.89	0.55		

(b) Subject 1 as testing data.

	FBM [100 fps]				
UF	0	10	20	30	40
5	0.89	0.65	0.55	0.52	0.49
10	0.89	0.65	0.55	0.54	0.50
15	0.89	0.64	0.55	0.54	0.50
20	0.87	0.64	0.55	0.55	0.51

Table 5. Classification accuracy on left-right jump movement.

(a) Subject 7 as testing data.

	FBM [100 fps]				
UF	0	10	20	30	40
5	0.84	0.69	0.76	0.63	0.52
10	0.82	0.75	0.66	0.58	0.48
15	0.82	0.73	0.79	0.55	0.52
20	0.84	0.83	0.85	0.57	0.51
25	0.83	0.89	0.73	0.58	
30	0.85	0.85	0.79	0.58	
35	0.87	0.91	0.85	0.53	
40	0.86	0.87	0.77	0.54	

(b) Subject 1 as testing data.

	FBM [100 fps]				
UF	0	10	20	30	40
5	0.96	0.85	0.70	0.51	0.49
10	0.94	0.86	0.67	0.51	0.46
15	0.94	0.80	0.64	0.52	0.46
20	0.93	0.80	0.64	0.50	0.45
25	0.90	0.78	0.66	0.48	0.45
30	0.90	0.78	0.64	0.51	0.45
35	0.89	0.79	0.64	0.54	0.45
40	0.90	0.77	0.63	0.51	0.43

Table 6. Comparison of beginning of the main motions in trainings data. When the information from Subject 1 was used as testing data, Subjects 4–7 were used to train the proposed network. When Subject 7 was used as testing data, Subjects 1–4 were used for training.

No.	Main motion, average of Subjects 4-7 (Subject 1 as testing data source) (Mean±SD) [ms], $n = 480$	Main motion, average of Subjects 1–4 (Subject 7 as testing data source) (Mean±SD) [ms], $n = 480$	SD (f-test) $\alpha = 0.01$	Mean (Welch's t-test) $\alpha = 0.01$
Button touch	361±51	404±70	Significant difference	Significant difference
Kick	584±80	637±37	Significant difference	Significant difference
Rotation jump	669±148	704±126	Significant difference	Significant difference
Left-right jump	732±65	740±81	Significant difference	$p = 0.087$

seemed to be almost no preliminary motion, and it was difficult to classify the button-touch movement before a hand started to move.

For the kick movement, more than 90% of actions could be predicted by using body movement 20 frames back from a starting point (FBM = 0 and UF = 20) for both subjects. In contrast, prediction based on information up to 0.1 s before the starting point varied depending on the subject; for Subject 1, approximately 80% of the actions could be predicted using 15–20 frames. Because it was difficult to classify left or right jumping movements with inputs more than 0.2 s before the main motion (FBM $>$ 20), the preliminary motion was considered to occur within 0.2 s before the starting point, and we found that 15–20 frames of input enabled reasonably accurate prediction.

For the rotation jump, one subject could be predicted 0.1 s before the starting point and the other could not. The preliminary motion was thought to be within 20 frames before the start of the main motion.

The left-right jumping action showed the largest preliminary motion. Using body movement 0.2 s before the starting point, 70% of these actions could be predicted, and some conditions were 85% predictable. Preliminary motion appeared in some cases up to 40 frames before the main motion started. Five frames of input were sufficient to predict left or right jumps.

From these results, it was confirmed that the duration of the time defined as preliminary motions in Section 4.1 was correlated with the predictable duration of the movement. In other words, the movements of the other body parts that occurred before the beginning of the main motion observed in Section 4.1 were considered to be preliminary motions with sufficient information to predict which way the subject would move. In addition, although there were differences depending on movement type, prediction accuracy was higher for UFs around 15–30, suggesting that an input length of 0.15–0.3 s could be effective for applications such as real-time estimation.

In addition, in order to examine the effect of the differences in training data, we summarized the means and standard deviations of the training data for Subjects 1 and 7 in **Table 6**. In most cases, the variances were not equal, and the means were different; therefore, the training and validation data may also be considered a factor for the difference in estimation accuracy.

Table 7. Classification accuracy on the button touch movement, excluding the input of right arm. Subject 7 as testing data.

	FBM [100 fps]			
UF	−20	−10	0	10
5	0.58	0.48	0.48	0.48
10	0.53	0.48	0.46	0.48
15	0.47	0.47	0.50	0.48
20	0.47	0.47	0.48	0.52

5. Discussion

5.1. Arm Movement Prediction

To investigate the influence of body parts that were not important in predicting the main motions, we also performed left-right classification in a case where important skeletal regions were excluded from the input skeletal coordinates. In the case of the button touch movement, the input per frame was 17×3 points, excluding the 4 point coordinates of the right arm, and for the other actions, the input per frame was 13×3 points excluding the eight points of both legs.

As a result, the accuracy of the kick, rotation, and translation jump movements did not change significantly, but the accuracy of the button touch changed significantly (**Table 7**). Therefore, for the button touch movement, it is considered that the left-right decision for FBM $= -20$ was made using the actual movement of the right arm after the main movement started. No other body parts contributed to the estimation of motion for this action.

As for the reasons why preliminary motions did not appear in the button touch movement as they did in others, the following can be considered.

- The physical torque required to move an arm is relatively small and requires relatively little support from other parts of the body.
- The kicking, rotating, and left-right jumping movements involve a need to balance the body.

The first factor could be investigated by making the same measurement with a weight on the arm or holding a stick to change the physical moment, and classifying the left and right in that case. As for the second factor, if the accuracy of classification were decreased in a balanced situation (e.g., kicking from a seated position), it could be confirmed that the contribution of balance maintenance observations to the accuracy of left-right classification is significant.

5.2. Network Structure and Prediction

In general, one NN failing to predict does not mean that prediction is impossible with that dataset, because other NNs or methods may be able to make a correct prediction. In this case, however, we were able to identify the period of time when a preliminary motion occurred from the motion capture data, and there was a good correlation between that period of time and the time span predictable by NN. Therefore, it is reasonable to say that prediction cannot be performed well using data lacking preliminary motion.

Currently, room for improvement remains because prediction is not perfect in areas where preliminary movements are observable. However, in many situations where prediction has not been possible, preliminary movements have not been observed, and it is thus highly unlikely that prediction will be possible even if a more advanced NN structure were devised.

On the other hand, it is possible that different types of information, such as mutual information, could be used to extract behavior-related information earlier than a NN can. This is considered as a future challenge.

5.3. Identifying a Motion Start Moment

In this study, we have discussed whether or not there was a preliminary motion in some particular movements and whether or not the preliminary motion could be used for prediction. The length of these preliminary movements varied significantly depending on the definition of the motion start moment. We defined the motion start moment for limb movements as the moment when the part of the body we instructed to move started to move. In jumping actions, because the intended motion is a whole-body movement and it is not possible to define a specific part of the body to be moved, it is difficult to define a motion start moment. As mentioned above, we determined the timing from "the moment when the center of gravity began to move upward."

If the definition of the motion start moment was changed, the judgment of whether the prediction was possible would be accordingly altered. For example, the motion start moment may be defined as the moment when the brain processes the directional instruction on the screen and decides to move, or as the moment when myoelectricity is observed in the target muscle. In such a case, the motion start moment would be set before the body movement started, and thus there would be no preliminary movement by definition. In this paper, we discuss whether there was a preliminary motion in terms of a motion start point defined within the range that can be judged by motion capture body movement measurement.

5.4. Prediction in Robotics

Based on our experimental results, we expect it to be difficult to predict fine gestures and eliminate delays, such as in telexistence, because subtle hand movements that do not disrupt the body's balance cannot be predicted.

One possibility is to utilize a "context of action." For example, Mao et al. [23] improved prediction performance by more than 0.5 s by including motion smoothness as a feature. If we consider the use of more medium- to long-term information, it may be possible to predict further ahead. Algorithms that take into account the context of an action, such as regressive neural networks, have the potential to improve prediction performance and have potential applications in robotics.

This study did not take such "context of action" into account, and assumed that only the preliminary motions associated with the main motions were directly used for prediction. We showed the theoretical limits of prediction without considering motion context. The results demonstrate that it is possible to estimate whole-body movement actions, such as jumping. This can be used in human-robot interaction spaces to avoid collisions by predicting sudden human movements.

5.5. Prediction in Group Behavior

In this study, it was shown that movements that involve shifting the center of gravity could be predicted. It is thought that humans can predict movements that have such preliminary motions and make use of this ability in human groups [24]. For example, people decide their own behavior by predicting the actions of others, such as walking in crowds. In the case of animals that move in a unified manner as a group, it may be meaningful to consider the effects of an individual's predicted movements of other individuals on their current movements and to consider their interactions. Although this study only focused on human movements, analysis and application to other animals may be possible and remain as challenging future issues.

6. Conclusion

In this study, we measured arm movements, leg movements, and whole-body movements at 100 fps to determine the motion starting point of a single motion from a static state. We also performed left and right classification using neural networks by changing the number of frames and input timing of the input skeletal coordinates. In this way, we identified sequences in which the necessary preliminary motions appeared before the main motions were performed. No preliminary motion was seen in the case of the button touch movement, but for the kicking and rotation jumping movements a preliminary motion was seen 0.2 s before the main movement started, whereas for the left-right jumping action, the preliminary motion began 0.3 s before the main movement. Based on this experimental data, it was confirmed that prediction by the neural network was possible for the actual movement in proportion to the time when a preliminary motion appeared. In other words, hand movements could not be predicted in advance, while kicks and jumps could be predicted from 0.1 s before, and left and right jumps from 0.2 s before. In the case of predicting other body movements in a forward propagation network, input data of approximately 0.15–0.3 s in length, 0.1–0.2 s before the start of the movement to be predicted are expected to be suitable.

These basic measurements of a person's onset of action and preliminary motions are basic information that helps enable an understanding of the behavior of people in groups. This is because people predict the movements of others based on their observations of them and change their own behavior accordingly, and the aggregation of these changes determines the behavior of people as a group. For example, one could simulate changes in the behavior of a group using models representing varying predictions times ahead for each individual, for instance 0.1 s ahead compared to 0.5 s. We believe that this basic information will prove useful to future clarification of the group behavior of humans and other living creatures.

Acknowledgements

This study was supported by JST PRESTO 17939983.

References:

[1] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," Robotics and Autonomous Systems, Vol.42, No.3-4, pp. 143-166, 2003.

[2] C. Breazeal, "Robot in society: Friend or appliance?," Proc. of the Agents99 Workshop on Emotion-Based Agent Architectures, pp. 18-26, 1999.

[3] N. Imaoka, K. Kitazawa, M. Kamezaki, S. Sugano, and T. Ando, "Autonomous Mobile Robot Moving Through Static Crowd: Arm with One-DoF and Hand with Involute Shape to Maneuver Human Position," J. Robot. Mechatron., Vol.32, No.1, pp. 59-67, 2020.

[4] Y. H. Weng, C. H. Chen, and C. T. Sun, "Toward the human-robot co-existence society: On safety intelligence for next generation robots," Int. J. of Social Robotics, Vol.1, No.4, pp. 267-282, 2009.

[5] G. K. Still, "Crowd dynamics," Ph.D. Dissertation, University of Warwick, 2000.

[6] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2891-2900, 2017.

[7] T. Vicsek and A. Zafeiris, "Collective motion," Physics Reports, Vol.517, No.3-4, pp. 71-140, 2012.

[8] M. Ballerini et al., "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," Proc. of the National Academy of Sciences, Vol.105, No.4, pp. 1232-1237, 2008.

[9] R. Lukeman, Y. X. Li, and L. Edelstein-Keshet, "Inferring individual rules from collective behavior," Proc. of the National Academy of Sciences, Vol.107, No.28, pp. 12576-12580, 2010.

[10] K. Fragkiadaki, S. Levine, P. Felsen, and J. Malik, "Recurrent network models for human dynamics," Proc. of the IEEE Int. Conf. on Computer Vision, pp. 4346-4354, 2015.

[11] H. Chiu, E. Adeli, B. Wang, D. Huang, and J. C. Niebles, "Action-agnostic human pose forecasting," Proc. of the 2019 IEEE Winter Conf. on Applications of Computer Vision (WACV), pp. 1423-1432, 2019.

[12] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.36, No.7, pp. 1325-1339, 2014.

[13] E. Barsoum, J. Kender, and Z. Liu, "HP-GAN: Probabilistic 3D human motion prediction via GAN," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops, pp. 1418-1427, 2018.

[14] Y. Chao, J. Yang, B. Price, S. Cohen, and J. Deng, "Forecasting human dynamics from static images," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 548-556, 2017.

[15] M. Bataineh, T. Marler, K. Abdel-Malek, and J. Arora, "Neural network for dynamic human motion prediction," Expert Systems with Applications, Vol.48, pp. 26-34, 2016.

[16] Y. Horiuchi, Y. Makino, and H. Shinoda, "Computational foresight: Forecasting human body motion in real-time for reducing delays in interactive system," Proc. of the 2017 ACM Int. Conf. on Interactive Surfaces and Spaces, pp. 312-317, 2017.

[17] S. Suda, Y. Makino, and H. Shinoda, "Prediction of Volleyball Trajectory Using Skeletal Motions of Setter Player," Proc. of the 10th Augmented Human Int. Conf. 2019, pp. 1-8, 2019.

[18] E. Wu and H. Koike, "FuturePong: Real-time Table Tennis Trajectory Forecasting using Pose Prediction Network," Proc. of the Extended Abstracts of the 2020 CHI Conf. on Human Factors in Computing Systems, pp. 1-8, 2020.

[19] E. Wu and H. Koike, "FuturePose-Mixed Reality Martial Arts Training Using Real-Time 3D Human Pose Forecasting With a RGB Camera," Proc. of the 2019 IEEE Winter Conf. on Applications of Computer Vision (WACV), pp. 1384-1392, 2019.

[20] A. Namiki and F. Takahashi, "Motion Generation for a Sword-Fighting Robot Based on Quick Detection of Opposite Player's Initial Motions," J. Robot. Mechatron., Vol.27, No.5, pp. 543-55, 2015.

[21] S. Tachi, K. Tanie, K. Komoriya, and M. Kaneko, "Tele-existence (I): Design and evaluation of a visual display with sensation of presence," A. Morecki, G. Bianchi, and K. Kędzior (Eds.), "Theory and Practice of Robots and Manipulators," MIT Press, pp. 245-254, 1985.

[22] T. Kurai, Y. Shioi, Y. Makino, and H. Shinoda, "Temporal Conditions Suitable for Predicting Human Motion in Walking," Proc. of the 2019 IEEE Int. Conf. on Systems, Man and Cybernetics (SMC), pp. 2986-2991, 2019.

[23] W. Mao, M. Liu, M. Salzmann, and H. Li, "Learning trajectory dependencies for human motion prediction," Proc. of the IEEE Int. Conf. on Computer Vision, pp. 9489-9497, 2019.

[24] M. Argyle, "Non-verbal communication in human social interaction," Cambridge University Press, 1972.

Name:
Chaoshun Xu

Affiliation:
Graduate School of Information Science and Technology, The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Brief Biographical History:
2019- Graduate School of Information Science and Technology, The University of Tokyo

Name:
Masahiro Fujiwara

Affiliation:
Project Assistant Professor, Graduate School of Frontier Sciences, The University of Tokyo

Address:
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan

Brief Biographical History:
2015 Received Ph.D. in Information Science and Technology from The University of Tokyo
2015- Toshiba Corporation
2018- Project Assistant Professor, Graduate School of Frontier Sciences, The University of Tokyo

Main Works:
• M. Fujiwara and H. Shinoda, "Remote Measurement of Surface Compliance Distribution Using Convergent Ultrasound," Trans. of the Society of Instrument and Control Engineers, Vol.49, No.4, pp. 455-460, 2013 (in Japanese).

Membership in Academic Societies:
• The Society of Instrument and Control Engineers (SICE)

Name:
Yasutoshi Makino

Affiliation:
Associate Professor, Graduate School of Frontier Sciences, The University of Tokyo
JST PRESTO

Address:
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan
7 Gobancho, Chiyoda-ku, Tokyo 102-0076, Japan

Brief Biographical History:
2007 Received Ph.D. in Information Science and Technology from The University of Tokyo
2008-2009 JSPS Research Fellowship for Young Scientists / Project Researcher, The University of Tokyo
2009-2013 Project Assistant Professor, Keio University
2013-2017 Lecturer, The University of Tokyo
2017- Associate Professor, The University of Tokyo

Main Works:
• Y. Makino, Y. Furuyama, S. Inoue, and H. Shinoda, "HaptoClone (Haptic-Optical Clone) for Mutual Tele-Environment by Real-time 3D Image Transfer with Midair Force Feedback," Proc. of the 2016 CHI Conf. on Human Factors in Computing Systems, pp. 1980-1990, 2016.

Membership in Academic Societies:
• The Society of Instrument and Control Engineers (SICE)
• The Virtual Reality Society of Japan (VRSJ)
• Information Processing Society of Japan (IPSJ)

Name:
Hiroyuki Shinoda

Affiliation:
Professor, Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, The University of Tokyo

Address:
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8561, Japan

Brief Biographical History:
1995- Lecturer, Tokyo University of Agriculture and Technology
1997- Associate Professor, Tokyo University of Agriculture and Technology
1999- Visiting Researcher, University of California, Berkeley
2000- Associate Professor, The University of Tokyo
2012- Professor, The University of Tokyo

Main Works:
• Information physics, haptics, mid-air haptics, sensor systems and devices, two-dimensional communication, human interfaces, and their application systems

Membership in Academic Societies:
• The Society of Instrument and Control Engineers (SICE)
• The Japan Society of Mechanical Engineers (JSME)
• The Institute of Electrical Engineers of Japan (IEEJ)
• The Robotics Society of Japan (RSJ)
• The Virtual Reality Society of Japan (VRSJ)
• The Institute of Electrical and Electronics Engineers (IEEE)
• Association for Computing Machinery (ACM)